

# An Extended Hierarchical Task Analysis for Error Prediction in Medical Devices

Phillip H. Chung<sup>1,2</sup>, Jiajie Zhang, PhD<sup>1</sup>, and Todd R. Johnson, PhD<sup>1</sup>, and Vimla L. Patel, PhD<sup>3</sup>

<sup>1</sup>School of Health Information Sciences, University of Texas, Houston TX

<sup>2</sup>Department of Psychology, Rice University, Houston TX

<sup>3</sup>Department of Medical Informatics, Columbia University, New York NY

## ABSTRACT

*This paper introduces an extended hierarchical task analysis (HTA) methodology devised to evaluate and compare user interfaces on volumetric infusion pumps. The pumps were studied along the dimensions of overall usability and propensity for generating human error. With HTA as our framework, we analyzed six pumps on a variety of common tasks using Norman's Action theory. The introduced method of evaluation divides the problem space between the external world of the device interface and the user's internal cognitive world, allowing for predictions of potential user errors at the human-device level. In this paper, one detailed analysis is provided as an example, comparing two different pumps on two separate tasks. The results demonstrate the inherent variation, often the cause of usage errors, found with infusion pumps being used in hospitals today. The reported methodology is a useful tool for evaluating human performance and predicting potential user errors with infusion pumps and other simple medical devices.*

## INTRODUCTION

**Predicting Human Error:** The study of human error and error prediction is one that has long incubated within its parent fields of human factors and systems engineering. It has given rise to any number of formal methodologies, such as fault tree analysis and failure modes and effects analysis<sup>1</sup>. All share the common goal of helping analysts predict and analyze the types of errors that can occur in any situation of human and machine interaction.

More recently interest has spread to cognitive theory-based methods of task description and error prediction that focus on people and how they think and behave<sup>2</sup>. These aim to explain failures on the human side of the interaction using theory and findings from cognitive psychology. Rather than treating human error as a product of some invariable stochastic function, their purpose is to pinpoint specific causes and mechanisms.

For instance, GOMS analysis<sup>2</sup> breaks down tasks into Goals, Operators, Methods, and Selection rules to describe task procedures in a way that accounts for

human cognition and performance. Since it describes the procedural knowledge involved in a task, it can be used to model the human side of any instance of human-computer interaction and help identify potential areas of operator error.

It is critical to apply these types of techniques to medical devices because the underlying problem with them is often cognitive<sup>3</sup>. This is increasingly becoming true as medical devices become progressively computerized. Hence, effective design and redesign efforts to promote comfortable, effective, and safe use can only be done with consideration of human cognition.

**Human Error in Medicine:** Since the release of the Institute of Medicine medical error report in 1999<sup>4</sup>, human errors in medicine and patient safety have become of great concern in the medical field. Like other landmark events in the history of human factors, such as the nuclear accident at Three-mile Island, this occurrence and others have boosted awareness of human factors within this domain. Device level incidents stemming from poorly designed interfaces have become particularly problematic with advances in technology. The literature has long suggested that the number of injuries resulting from these types of problems far exceed that of injuries due to device failures<sup>5</sup>.

Today, most medical device manufacturer's websites and advertising claim that human factors are of high priority in their designs. Nevertheless, a simple search on the FDA Manufacturer and User Facility Device Experience Database (MAUDE)<sup>6</sup> for most common medical devices will bring up any number of reported "usage" problems. Considering the absence of a widespread, formalized means for predicting human errors at the human-device level and evaluating interfaces for error-generating propensity, this is not surprising. Should such a method exist, manufacturers would have greater ability to accurately test and design their devices for safe use. Moreover, purchasers concerned about safety would be better equipped to discern among models for use at their respective facilities.

While the prospect of uncovering an infallible method for predicting human error given an interface

and task still seems distant, the present state of theory and research on human error in the fields of human-computer interaction (HCI) and cognitive science seems sufficient to provide improvement over currently techniques. This will prove invaluable if successfully applied in the medical arena where human lives are consistently at stake. This work represents one such effort.

**Volumetric Infusion Pumps:** One pervasive device in medicine that has led to numerous medical error incidents is the innocuous volumetric infusion pump. Used for the controlled delivery of medication, these devices greatly advanced the state of intravenous therapy. Nevertheless, major problems with them have been revealed publicly<sup>7</sup>.

From the standpoint of pure research, the volumetric infusion pump provides a relatively simple benchmark medical device on which human error treatments and theory may be developed and tested. Having grown increasingly computerized and complex over the years, they have fallen into the domain of human-computer interaction (HCI) and cognitive engineering. It now seems more sensible than ever to apply existing knowledge from HCI and cognitive science pertaining to interface design in the medical domain.

## METHODS

**Hierarchical Task Analysis:** Hierarchical task analysis (HTA)<sup>1</sup>, one of the most widely used forms of task analysis, involves describing a task as a hierarchy of tasks and subtasks. The initial step in our own methodology was to construct HTAs of three typical infusion tasks for the pumps under study. Our approach is similar to that of Rogers and colleagues<sup>8</sup> who applied this basic technique in the evaluation of a simple medical device.

As Reason<sup>9,10</sup> recommends, a task analysis is critical because it provides the necessary information for the analysis and prediction of operator error (e.g., goals, tasks, and subtasks). It aptly represents the transitional flow of the interaction in a format useful to the evaluator (in plans of interaction). This information may then be iteratively used to extend evaluation to account for both distributed representations<sup>3</sup> and what Reason calls error “affordances”<sup>9,10</sup> or error-shaping factors predefined by psychological research (e.g., steps with high short term memory demands).

Decomposing task structures into reoccurring “components” or subtasks reveals identifiable patterns. We can highlight them simply on a spreadsheet using colors, permitting visual comparisons across different pump models. Longer procedures, intuitively, have been shown to decrease usability and increase potential for error<sup>8</sup>.

**Multiple Tasks:** It is unlikely that a single programming task will utilize all commonly used features on a pump. Hence, conducting task analyses on each pump for multiple tasks allows for a more thorough evaluation. Steps or interface artifacts that are particularly likely to elicit user errors will be identified as they reappear in the evaluation across tasks. By tabulating and comparing the number of external representations, error affordances, etc., we can expand our basis for prediction. Finally, collapsing the data provides a single metric of error propensity for general comparison across pumps. To allow for stronger comparisons, similar tasks are used both on the single and multiple channel pumps.

**Norman’s Action Cycle:** In order to bring the relevant cognitive processes into consideration, we conduct our analysis following Norman’s Action cycle<sup>11</sup>. This widely cited theory decomposes interaction between human and computer into seven basic stages of user activity (Figure 1).

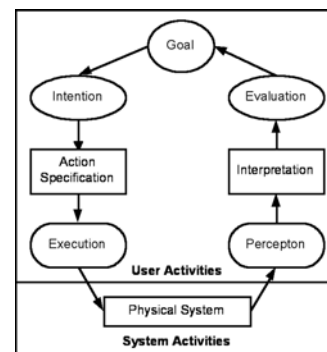


Figure 1. Norman’s Action Cycle

To reduce the possibility of human error, Norman recommends that the Gulf of Evaluation (right side), or the gap between Perception and Evaluation, and the Gulf of Execution (left side), or the gap between Intention and Action, be made as narrow as possible. Direct interaction is the ultimate goal. Hence, the device should provide quick and informative feedback and heavily nested functions should be avoided as much as possible. Any necessary information for operation should be displayed intelligently on the device itself.

**Distributed Representations:** Examining each discrete step in our task analysis along Norman's seven stages of action requires the evaluator to concomitantly identify the appropriate internal and external representations involved. According to Zhang and Norman<sup>12</sup>, this will establish the cognitive task space of the human-device interaction. Following the Action theory, external representations, where the information is present on the device itself rather than in the head of the operator, are thought to drastically reduce or eliminate human error. Internal representations depend on cognitive processes for retrieval, whereas external representations rely on more efficient perceptual processes, hence mitigating cognitive load. Of course, this is contingent upon the correct mapping of functions on the interface with the internal knowledge of the user.

For accurate evaluation of these representations, it is necessary to consider all information presented by the device. This includes auditory warnings, text messages, LEDs, physical controls, etc., which determine the external problem space and provide context for the evaluator. Research shows that the specific wording and presentation of labels and

feedback messages have great impact on the actions users decide to take<sup>13</sup>. Listing all state and transitional information makes it possible for the evaluator to identify deficiencies in the external problem space.

**Output Presentation:** Presenting the final output of these theories and methodologies in a compact tabular form (Figure 2) allows the evaluator to visualize and predict where possible errors may occur, rate their severity, estimate relative task completion times, and, most importantly, contrast across different devices. This type of detailed cognitive task analysis can be useful for pump manufacturers, hospitals, and other decision makers to assess and compare different pump models taking into account patient safety-critical information.

Evaluation of our analyses reveals a great deal of variability in task space constraints as well as error affordances between the six pumps under study, predicting that some will be more error prone and difficult to use than others. For this paper we will contrast findings from two different pumps and compare our relative predictions to FDA error reports from MAUDE.

Number	Goal/Method	GOMS Operators	Physical State	Previous Screen	Current Screen	Auditory Warnings	Softkey 1
<b>1.00</b>	<b>Prepare Device</b>						
1.10	Plug in	Insert plug	Plugged in/off/unloaded				
1.20	Make sure Lockout is off	Verify Lockout status					
1.30	Turn device on	Press 'Power'	Plugged in/on/unloaded		Instrument information; Standard display	Audio test "beep"	STNDBY
<b>2.00</b>	<b>Prepare Set</b>						
2.10	Move AccuSlide flow regulator clamp down until it clicks closed	Slide clamp					
2.20	Spike the fluid container.	Spike container.	Plugged in/on/unloaded	Standard display	Standard display		STNDBY
2.30	Fill drip chamber to fill line	Wait for chamber to fill	Plugged in/on/unloaded	Standard display	Standard display		STNDBY
2.40	Open AccuSlide clamp to prime set	Open clamp; wait	Plugged in/on/unloaded	Standard display	Standard display		STNDBY
2.50	Close AccuSlide clamp when priming is complete	Close clamp	Plugged in/on/unloaded	Standard display	Standard display		STNDBY

**Figure 2. Tabular output (sample)**

## RESULTS

**Sample:** To make the methodology and process more concrete, a sample analysis will be presented in detail using two unnamed single channel pumps X and Y. Pump X is a widely used single channel volumetric infusion pump from a large, longstanding medical device manufacturer. Well recognized and touted for its human-centered design, the pump brings with it several human factors improvements over the previous technology. Nevertheless, a quick search on the FDA database proves that it is not problem-free.

Using pump X and a competing pump Y from another large manufacturer (>30% U.S. market

share), we will demonstrate how this methodology is applied. Both single channel pumps will be analyzed on two different tasks to show their intrinsic differences.

**Procedure:** In this sample evaluation we will use two of the original three single channel tasks. Task 1 and 2 follow:

1. Orders: Infuse 1000 ml Normal Saline at 125 ml per hour.  
Goal: Verify fluid is flowing to patient.

2. Orders: Infuse 1000 ml of NS at 125 ml per hours as a continuous infusion. Give 1gm Ceftriaxime (Fortaz) IVPB q 8 hr. Administer over 30 minutes.  
Goal: Verify fluid is flowing to patient.

Starting from a given order and goal, we construct our initial HTAs. Operations vary depending on the particular pump, but major subtasks, such as the programming of an infusion rate or the loading of tubing remain constant. Using expert knowledge and the manufacturer's user manuals, tasks are decomposed into their constituent subtasks and operations. When finalized, this leaves us with a basic framework for each pump and task.

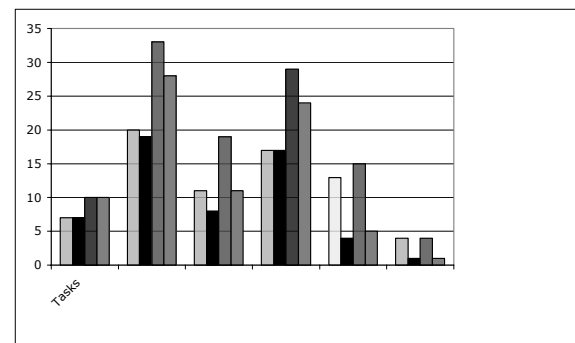
State information is next added aside each step in the table and indicates, most generally, the status of the device. For example, a given state might indicate that the device is powered on. More detailed information is provided under separate columns for auditory cues, display messages, "softkey" information, etc. Previous state and display information dictate the context the user is given as they proceed to the next step. This information is used to decide what must be done on the subsequent screen, and thus has a great effect on the user's behavior.

Using the completed state, display, auditory, and key information columns, we fill in a column for external representations<sup>12</sup>. For the internal representations we use a GOMS type analysis that answers the question of what declarative knowledge or mental operator is needed by a user to complete a particular subtask.

Finally, from these two columns we follow the seven steps of the Action cycle to deduce where possible errors might arise (affordances). Going through the steps (e.g., Perception) for each subgoal (e.g., turn device on), we check if the device provides sufficient external representations (e.g., Is there information on the device to indicate how to turn on the device?) to satisfy the needs and deficiencies of the internal rules or representations. A poorly determined external problem space suggests room for human error. Additionally, steps that fit one of Reason's predefined error provoking factors (affordances)<sup>9,10</sup> are also noted. A few of these appeared across both tasks, presenting a greater likelihood for error.

**Analysis:** Figure 3 shows a breakdown of the total number of steps, internal and external representations, and error affordances for pumps X and Y on Tasks 1 and 2, respectively. Conditional steps, or steps necessary only given some precondition were omitted.

The number of MAUDE usage-related errors reported during the 2001-2002 period has also been added under "Reports" for reference. These do not include mechanical failures.



**Figure 3. Task 1 and 2 totals**

Comparisons between the pumps are possible by simply counting the number of overall steps or operations needed<sup>8</sup>, as well as the number of affordances. Looking at Figure 3, pump X has nine (Task 1) and ten (Task 2) more affordances. This follows the trend of usage-related problems found for each pump on MAUDE (four for pump X versus one for pump Y) in 2001.

## DISCUSSION

Six pumps (three single channel and three multichannel) in all were analyzed using the method described. As with pumps X and Y, there was a large degree of variation on the final output of the analyses. While it may be argued that the pumps with fewer required operations also had fewer options, counting options was not the focus of this methodology. Undoubtedly, there will be some tradeoffs made in any engineering or purchasing decision. Moreover, more steps did not always translate to more affordances, as some pumps had extra steps for "forcing functions". These act to force the user to verify their settings or inputted values and thus actually contribute to overall safety.

Even pumps from the same manufacturer showed large differences in their task structures. This presents a potential problem for hospitals employing temporary nurses who may need to use different pump models with little or no training beforehand. In separate controlled experiments, several fourth year nursing students complained that they could not complete the tasks because, "they did not have much experience" with a certain pump. Many went on to incorrectly complete the programming tasks. This is in spite of being given a ten-minute demonstration of the pump functions prior to the test, as is often the case in practice.

## CONCLUSION

The methodology reported here will not predict every possible human error given an interface. The capricious nature of human behavior makes this nearly impossible. As is the shortcoming of all techniques in interface evaluation, it also offers no suggestions for redesign.

However, a careful evaluator should be able to make general predictions of relative error propensity between pumps and identify potential error inducing features and steps. While the technique remains dependent upon the skill of the evaluator, we believe it is a step in the right direction. It upholds the value of considering cognition when assessing the safety of a medical device through the employment of cognitive theory. And our methodology can be practical; as demonstrated, simply tabulating the external features of an interface side-by-side with the steps required allows quick contrasts across different pump models.

**Work in Progress:** To further validate and refine our methods, we will compare our analyses and predictions with data from controlled user studies. We are also working to streamline this process, to reduce dependency on the evaluator's skill level.

**Suggestions for Further Work:** These analyses provide all of the necessary information to produce a working computational model. Existing unified theories of cognition, such as ACT-R<sup>14</sup>, already provide a proven framework for predictive models that can account for human cognition and well describe interactive behavior and error<sup>15</sup>. Extending this type of approach to the medical domain could indeed prove beneficial, as safety and time critical applications abound.

## REFERENCES

1. Kirwan B, Ainsworth, LK. A Guide to Task Analysis. London: Taylor & Francis; 1992.
2. John BE, Kieras DE. The GOMS family of analysis techniques: Tools for design and evaluation. Carnegie Mellon University School of Computer Science Technical Report No. CMU-CS-94-181. 1994.
3. Zhang J, Patel, VL, Johnson TR. Medical error: Is the solution cognitive or medical? Journal of American Medical Informatics Association 2002;9(6):S75-S77.
4. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human. Washington, D.C.: National Academy Press; 1999.

5. Cooper JB, Newbower RS, Long CD, McPeck B. Preventable anesthesia mishaps: a study of human factors. Anesthesiology 1978;49(6):399-406.
6. U.S. Food and Drug Administration (FDA). Center for Devices and Radiological Health. MAUDE Database [Online] [accessed February 2003]. Available from URL <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMAUDE/search.cfm>
7. U.S. Food and Drug Administration (FDA). Department of Health and Human Services. Public Health Advisory: Avoiding Injuries from Rapid Drug or I.V. Fluid Administration Associated with I.V. Pumps and Rate Controller Devices. Issued on March 1, 1994.
8. Rogers WA, Mykityshyn AL, Campbell RH, Fisk AD. Only 3 easy steps? User-centered analysis of a "simple" medical device. Ergonomics in Design 2001;9:6-14.
9. Reason J. Combating omission errors through task analysis and good reminders. Quality and Safety in Healthcare 2002;11:40-44.
10. Reason J. Managing the Risks of Organizational Accidents. Aldershot: Ashgate Publishing Company; 1997.
11. Norman DA, 1986. Cognitive Engineering. In Norman DA, Draper SW, editors. User Centered System Design: New Perspectives on Human-Computer Interaction. Hillsdale, NJ: Lawrence-Erlbaum; 1986. pp 31-61.
12. Zhang J, Norman DA. Representations in distributed cognitive tasks. Cognitive Science 1994;18:87-122.
13. Polson PG, Lewis CH. Theory-based design for easily learned interfaces. Human-Computer Interaction 1990;5:191-220.
14. Anderson JR, Lebière C. The atomic components of thought. Mahway, NJ: Erlbaum; 1998.
15. Lebière C, Anderson JR, Reder LM. Error modeling in the ACT-R production system. Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society; 1994; Hillsdale, NJ: Erlbaum; 1994. p. 555-559

## ACKNOWLEDGEMENTS

The project/effort depicted was sponsored by the Dept of Army under Cooperative Agreement #DAMD17-97-2-7016. The content of the information does not necessarily reflect the position/policy of the gov. or NMTB, and no official endorsement should be inferred. It was also in part supported by grant number P01HS11544 from the Agency for Healthcare Research and Quality.